

WHAT IS CLAIMED IS:

1. A method for predicting at least part of a binding site in a target protein, wherein said binding site can be bound by a molecule, said method comprising:

(a) comparing, for each of a plurality of cross-reactive proteins, each of a first plurality of amino acid sequences in a region of said target protein with each of a second plurality of amino acid sequences in a region of said cross-reactive protein, wherein each said cross-reactive protein can be bound by said molecule;

(b) identifying an amino acid sequence in said first plurality of amino acid sequences that exhibits the highest average sequence homology score, said average score being based upon the sequence homologies to an amino acid sequence in each of said second plurality of amino acid sequences in regions of said cross-reactive proteins, wherein said identified amino acid sequence in said first plurality of amino acid sequences is predicted to be said at least part of a binding site in said target protein.

2. A method for predicting at least part of a binding site in a target protein, wherein said binding site can be bound by a molecule, said method comprising:

(a) evaluating the degree of homology between each n-amino acid window of a plurality of n-amino acid windows of the target protein with each n-amino acid window of a plurality of n-amino acid windows of a first cross-reactive protein of a plurality of cross-reactive proteins, wherein (i) each cross-reactive protein in the plurality of cross-reactive proteins can be bound by the molecule, and (ii) n is between 6 and 25;

(b) performing step (a) for each cross-reactive protein in the plurality of cross-reactive proteins;

(c) identifying, for each n-amino acid window in the target protein, the highest degree of sequence homology with an n-amino acid window in a cross-reactive protein for each cross-reactive protein;

(d) identifying the n-amino acid window(s) in the target protein that have the highest average of the highest degrees of sequence homologies identified in step (c), wherein said identified n-amino acid window(s) comprises at least part of the binding site(s) in the target protein.

3. A method for predicting at least part of a binding site in a target protein, wherein said binding site can be bound by a molecule, said method comprising:

- (a) comparing each n-amino acid window in a plurality of n-amino acid windows of the target protein with each n-amino acid window in a plurality of n-amino acid windows of a first cross-reactive protein of a plurality of cross-reactive proteins, wherein (i) each cross-reactive protein in the plurality of cross-reactive proteins can be bound by the molecule, and (ii) n is between 6 and 25;
- (b) assigning a score for each n-amino acid window comparison of step (a), wherein the score reflects the degree of sequence homology between the two n-amino acid windows compared;
- (c) performing steps (a) and (b) for each cross-reactive protein in the plurality of cross-reactive proteins;
- (d) identifying the highest scores assigned in step (b) of each n-amino acid window in the target protein for each cross-reactive protein; and
- (e) identifying the n-amino acid window(s) in the target protein that have the highest average score(s), wherein said identified n-amino acid window(s) comprises at least part of the binding site(s) in the target protein.

4. The method of claim 1, 2, or 3, wherein the binding site is an epitope and the molecule is an antibody.

5. The method of claim 1, 2 or 3, wherein the degree of sequence homology reflects the degree of sequence identity.

6. The method of claim 1, 2 or 3, wherein the degree of sequence homology reflects the degree of sequence similarity.

7. The method of claim 1, wherein the first plurality of amino acid sequences comprises successive overlapping amino acid sequences spanning said region of said target protein.

8. The method of claim 1, wherein said plurality of amino acid sequences of each said cross-reactive protein comprises successive overlapping amino acid sequences spanning said region of said cross-reactive protein.

9. The method of claim 7, wherein said successive overlapping amino acid sequences span said region of said target protein at an amino acid interval of 1 amino acid.
10. The method of claim 8, wherein said successive overlapping amino acid sequences span said region of said cross-reacting protein at a amino acid interval of 1 amino acid.
- 5 11. The method of claim 2 or 3, wherein the plurality of n-amino acid windows in the target protein comprises successive, overlapping amino acid sequences spanning a region of the target protein.
12. The method of claim 2 or 3, wherein the plurality of n-amino acid windows in each cross-reactive protein comprises successive overlapping amino acid sequences spanning a
10 region of the cross-reactive protein.
13. The method of claim 11, wherein said successive overlapping amino acid sequences span said region of said target protein at an amino acid interval of 1 amino acid.
14. The method of claim 12, wherein said successive overlapping amino acid sequence span said region of said cross-reactive protein at an amino acid interval of 1 amino acid.
- 15 15. The method of claim 1 or 11, wherein the region of the target protein has been identified as containing the binding site.
16. The method of claim 1 or 12, wherein the region of the cross-reactive protein has been identified as containing the binding site.
17. The method of claim 1 or 11, wherein the region of the target protein consists of the
20 entire contiguous amino acid sequence of the target protein.
18. The method of claim 1 or 12, wherein the region of the cross-reactive protein consists of the entire contiguous amino acid sequence of the cross-reactive protein.
19. The method of claim 1 or 11, wherein the region of the target protein has been identified as being on the surface of the folded target protein.
- 25 20. The method of claim 1 or 12, wherein the region of the cross-reactive protein has been identified as being on the surface of the folded cross-reactive protein.
21. The method of claim 1, 2, or 3, wherein the method is computer-implemented.

22. A computer system comprising a processor, and a memory coupled to said processor and encoding one or more programs, wherein said one or more programs cause the processor to carry out the method of any one of claims 1, 2, and 3.

5 23. A computer program product for use in conjunction with a computer having a processor and a memory connected to the processor, said computer program product comprising a computer readable storage medium having a computer program mechanism encoded thereon, wherein said computer program mechanism may be loaded into the memory of said computer and cause said computer to carry out the method of any one of claims 1, 2,
10 and 3.